

Study **Protocol**

Project title: The IBD BioResource

Full Title: The Inflammatory Bowel Disease BioResource: Progressing from Genetics to Function and Clinical Translation in Crohn's Disease & Ulcerative Colitis

Contents

1. Introduction

2. Aims of the Study

3. Background

3.1. Genetic prelude

3.2. Integrating successful current BioResources

3.3 Specifics of the IBD BioResource

4. Sample size for added value

5. Stage 1 activity - Recruitment to the IBD BioResource panel

5.1. Eligibility criteria

5.2. Route to recruitment

5.3. Data collection and storage

5.4. Samples

5.5. Reporting of results

5.6. Participant safety

6. IBD BioResource funding and Organisational Structure

7. Stage 2 activity - Research studies involving BioResource volunteers and/or their data

8. Appendix - detailed rationale for sample size

9. References

1. Introduction

The IBD BioResource is being set up by the UK IBD Genetics Consortium and the NIHR BioResource to collect biological samples and data from a cohort of over 25,000 individuals with Crohn's disease or ulcerative colitis. Inflammatory Bowel disease, or IBD, is the collective term for these conditions. There are 2 arms of recruitment:

1/ The Main cohort comprises individuals with established Crohn's Disease, Ulcerative Colitis or IBD type unspecified. Both clinical and self-reported phenotype data will be collected, alongside plasma, serum and DNA samples for genetic analysis, including whole Genome Sequencing.

2/ The Inception cohort aims to recruit a sub-set of 1,000 individuals newly diagnosed with IBD. More detailed samples including stool, biopsy tissue and whole blood for RNA and longitudinal follow-up of their medical course will be obtained from clinical records and from the subjects themselves.

This panel of IBD patients will be integrated within the existing cohort of the NIHR BioResource, with its National Coordinating Centre based in Cambridge. The NIHR BioResource provides researchers with groups of recallable volunteers and / or sample sets, tailor-made to the research question at hand, though not only for IBD research. The goal is to establish a sampling frame from which cohorts of patients can be selected on the basis of their genotype and/or phenotype to be invited for observational studies or interventional studies referred to as 'Stage 2' research activity. Invitation of patients for Stage 2 studies will be managed by the NIHR BioResource, and these studies will have to have been approved by the NIHR BioResource Scientific Advisory Committee (SAB) prior to inviting patients. In addition these DNA samples will be used in on-going genetic analyses by the UK IBD Genetics Consortium (UKIBDGC).

2. Aims of the study

The IBD BioResource aims to support studies looking at how genes and environmental factors influence disease and response to therapy. By gaining more information on the genes involved in Crohn's disease and ulcerative colitis and by understanding the differences in their function and integrating information regarding environmental influences including the microbiota we hope to gain insights regarding causal mechanisms, potential new therapies and treatment approaches for IBD to reduce the burden of disease, and begin to think about a cure.

Also, by working together with the NIHR BioResource, The IBD BioResource can benefit from being able to access a robust and tested system of re-call for volunteers to take part in medical research studies based on their genotype and/or phenotype.

3. Background

3.1 Genetic prelude

The last 10 years has seen major progress in identifying susceptibility genes for Crohn's disease and ulcerative colitis. Our UK IBD Genetics consortium, which is chaired by PI Miles Parkes, has played a key role in many of the pivotal studies which have been published in a sequence of reports in *Nature* and *Nature Genetics* [1-9].

The discovery of genes that influence risk of IBD and how the disease behaves, and interrogating the functions of these genes and their variants, has a very important role in helping to develop new interventions and treatments. As a step toward these goals it is essential to be able to correlate variation in genes with variations in characteristics such as disease behaviour, immune response and treatment response.

The classical genetic approaches narrow down areas of association to genes and single nucleotide polymorphisms or haplotypes, but causality can only be inferred if separate functional studies demonstrate an effect of the implicated variant on a key biological pathway known to lead to disease. This is relatively easy to do when rare mutations give rise to extreme phenotypes, but is much harder for less severe variants that are more common and which are neither necessary nor sufficient to cause disease. It is, however, these more common variants that may account for much of the inherited risk of common disease including IBD. Thus, as an adjunct to genetic association studies, it is important to establish studies in which the functional or therapeutic significance of genetic variants can be investigated. When a variant is very common, it may be possible to do this in an unselected population. However, for most variants this would be inefficient: it would be preferable to specifically select the population for functional and clinical studies based on carriage of the variant of interest.

3.2 Integrating successful current BioResources

In 2005, John Todd and colleagues established the NIHR Cambridge BioResource. This now has >17,000 healthy volunteers and patients who have

donated a sample of DNA and who are available for recall by genotype and/or phenotype. Members of our IBD research team have used this resource in our studies interrogating the function of IBD risk variants [10-11].

The IBD BioResource described in this application is designed to work in collaboration with the NIHR BioResource (with its National Coordinating Centre in Cambridge and a total of thirteen local BioResource centres based across England) by integrating a large cohort of patients with Crohn's disease or ulcerative colitis – such that we have available a large cohort of patients recallable by genotype (or disease phenotype) for future studies, markedly enhancing the efficiency by which we can undertake genotype-stratified studies. The NIHR BioResource will increase the size of its patient cohort and hence the total number of volunteers that can be invited to take part in research studies.

Recruitment of patients into the NIHR BioResource is covered in the NIHR BioResource (Research Tissue Bank) Protocol, IRAS ref: 220277. The IBD BioResource-associated patient information sheets (PIS and consent forms are based on those used by the NIHR BioResource – with some modifications made to incorporate the additional information required for a large cohort of subjects with IBD and also the genetics analysis which is integral to our proposal. The IBD BioResource makes this clear that the IBD BioResource sits within the NIHR BioResource.

In addition, the current proposal aims to encompass the samples, clinical information, genotyping and sequencing data that have been generated in the course of our earlier work, as part of the core activity of the constituent groups within the UK IBD Genetics Consortium (UKIBDGC). UKIBDGC and IBD Bioresource are closely integrated, with substantially overlapping investigators for both branches. The previous IBD genetics studies received ethics committee approval in earlier applications and substantial amendment proposals have been approved to respective REC's (for ref 05/Q0108/355 and 03/5/012) to allow roll-over of samples and data from these studies and collections into the proposed new IBD Bioresource study.

3.3 Specifics of the IBD BioResource

Unlike the main NIHR Bioresource cohort the current IBD Bioresource proposal includes our gene discovery programme as well as more detailed sampling from the inception cohort (see below) and many IBD-specific questions on

both the patient questionnaire and clinical phenotype form The genetic analysis is integral to the proposal and to the work of the UKIBDGC, the group of investigators of which sit at the core of the IBD Bioresource: it is more than simply the means by which subjects are identified for stage 2 studies which is the case for most subjects in the NIHR Bioresource programme. Although the ethics issues of consent for genotyping / sequencing for IBD BioResource are essentially the same as the ethics issues of consent for genotyping / sequencing for IBD gene discovery we feel that the emphasis is different. Furthermore, for our future IBD gene discovery programme we will aim to use DNA samples from the IBD BioResource rather than continuing recruitment of a separate cohort for UKIBDGC studies. It makes sense to consent such individuals using the single, uniform consent form attached to the current proposal rather than a 'new' IBD BioResource form and previous old UKIBDGC consent forms (which varied somewhat from centre to centre)

In this protocol, we describe the basis for the establishment of the IBD BioResource and summarise the principles under which subsequent 'Stage 2' research studies will be undertaken. However, as with the existing NIHR BioResource of healthy and patient volunteers, the ethical issues of the future functional and clinical studies will be particular to the study in question, and each of these Stage 2 research studies will require its own Ethics Committee application and permission prior to inviting BioResource participants. We are not seeking blanket approval for these studies with this application.

A key issue relate to notification of

- Pharmacogenetic findings – relating to variants which we and others are increasingly finding and which correlate with IBD treatment outcomes in terms of efficacy and adverse events
- Additional or secondary findings with potential clinical implications. Our policy in the IBD BioResource is directly aligned to the policy and protocol implemented at Genomics England and which has recently received REC approval (Ref 14/EE/1112). In formulating this we have been in discussion with Michael Parker, Professor of Bioethics at the University of Oxford and Director and Chair of the Ethics Advisory Group for Genomics England. In accordance with the Genomics England protocol we will offer potential volunteers the option of receiving feedback (or not) if they carry variants of potential medical significance in a specific list of 'pathogenic' genes which is monitored and updated by Genomics England.

We recognize that the genetic analysis that we are doing is in a research setting and not an accredited NHS genetics laboratory – and we have formulated our approach accordingly. In particular, information relating to pharmacogenetics will only be used in a research context and in the form of a clinician research survey; thus we have indicated on the patient information sheet that ‘

The IBD BioResource will not routinely feedback any genetic or biochemical research results obtained from your sample. As more genetic findings are identified, we will investigate their association with patterns of IBD and IBD treatment responses; this may include asking clinicians whether genetic information is useful to inform decisions regarding choice of treatment. In addition to the above, you will have a choice on the consent form to indicate if you would want to be informed in the rare event that you were found to be at increased risk of a genetic disease. This information may also be uncovered in the course of the research and might be relevant to your future health (for example, identifying an increased risk of a condition that might be prevented or treated early by the NHS). The updated list of the rare genetic conditions for which the option of this feedback is available on the Genomics England website at:

www.genomicsengland.co.uk/information-for-participants/findings/ (in ‘Additional or secondary findings’ section, expand the information box “if a participant agrees, we look for changes in the following genes”). If the genetic analysis indicates that you might be at increased risk of such conditions, and you choose to be informed about this, the provisional result of the research analysis will be handed back by the IBD BioResource team to your consultant or GP for discussion with you. A further sample would be taken for analysis within an accredited NHS diagnostics laboratory to confirm any “incidental finding” A health professional or genetic counsellor would then feedback to you on the final results.’

In addition to functional studies, the IBD BioResource will also prove of benefit in the study of the differential impact of lifestyle and other environmental factors including the microbiota by genotype. Observational studies have been established for investigating gene-lifestyle interaction through case-control studies nested with large cohorts (EPIC-Europe) and cross-sectional quantitative trait studies (e.g. Ely, Fenland). However, the causal inference from these observational studies would be greatly enhanced by evidence of differential response to intervention. An example is the demonstration of an apparent gene-diet interaction for the common Pro12Ala polymorphism in PPAR γ [12]. The case is strengthened by this combination of biologically relevant

environment-gene interaction plus a demonstrable link between the polymorphism and a discrete disease outcome [13]. Testing the hypothesis of dietary intervention and how genes and environmental factors interact would be greatly enhanced by being able to access a cohort of individuals of known genotype to study.

4. Sample size for added value

A very large sample size is important for a number of reasons as detailed in the appendix. In summary it will allow detailed evaluation of the contribution of low frequency and rare variants to disease susceptibility (for which sample sizes in the 10's of thousands are required); evaluation of the role of genetics in differences in disease course in IBD overall and in its constituent sub-phenotypes; an assessment of the role of gene-gene interaction (which for variants of modest frequency and low effect size, such as is typical in common disease, requires very large sample sizes) – and an assessment of gene-environment interactions (where similar arguments apply). Currently the IBD Bioresource has recruited ~20,000 subjects but there is definite scientific advantage in continuing recruitment. It is also true that many, many patients with IBD want to participate in this effort and we do not want to deny them that opportunity. One of the key funders for IBD BioResource is Crohn's and Colitis UK, the national patient support group, who see the IBD BioResource as a key vehicle to allow patients to participate in research.

5. Stage 1 activity – Recruitment to the IBD BioResource panel

5.1 Eligibility criteria

For both cohorts, volunteers must be 16 years of age or above. There is no exclusion on the basis of race, sex or disease status. Patients with mental and learning disabilities are included provided they are deemed to have capacity to provide informed consent. However, patients with significant cognitive impairment will be excluded from the study.

Specific eligibility criteria for the main cohort:

- Patients diagnosed with Crohn's disease encompassing: Crohn's Colitis, Crohn's pancolitis, Ileal Crohn's, Terminal Ileal Crohn's, Ileal-caecal Crohn's, Crohn's Ileitis, Crohn's jejuno-ileitis, Oral Crohn's and Oral Facial Granulomatosis

- Patients diagnosed with Ulcerative Colitis encompassing: Proctocolitis, Proctitis, Pancolitis, Total Colitis, Distal Colitis, Left-sided Colitis, Proctosigmoiditis
- IBDU: Patients with indeterminate colitis or IBD type unspecified (IBDU)

Specific eligibility criteria for the INCEPTION cohort:

- Patients newly diagnosed with Crohn's disease or ulcerative colitis or IBD type unspecified;

Or

- Suspected of having IBD based on the combination of symptoms, examination findings and baseline investigations (raised faecal calprotectin >200; and / or raised ESR or CRP).

5.2 Route to recruitment

Potential volunteers will be recruited through IBD clinics in participating hospitals across the UK. Some will be among the 19,000 current participants in our UK IBD Genetics Consortium (UKIBDGC) studies. Others may have participated in recent large IBD studies such as CONSTRUCT and TOPICC – these being particularly valuable due to the depth of phenotype data already ascertained, and their often severe disease phenotype. The majority, however, will be new to IBD research, recruited in hospitals not previously involved in the academic world but allowing these patients opportunity to contribute in a meaningful way to IBD research advances. Patient recruitment will be overseen by the NIHR BioResource and supported by CRN nurses (the study is funded by the MRC and Wellcome Trust, and is an NIHR portfolio study). Clinicians who recruited their patients to previous genetics studies led by the UKIBDGC will be contacted re the IBD Bioresource. Patients willing to join the IBD BioResource will be able to do so through the following routes:

- They will be able to join while attending their routine clinic appointment (most are on at least twice yearly follow-up in out-patients). Potential volunteers will be supplied with a letter of invitation and patient information sheet explaining the study requirements - either through the post ahead of their appointment time (we are, for example, in regular contact with UKIBDGC participants and will do a mailshot to them

regarding the opening of the IBD BioResource) or at the time of IBD clinic attendance. Where subjects have had sufficient time to consider involvement, and had the chance to discuss this if they wish (IBD clinics in participating hospitals will have a research nurse / CRN nurse in attendance to answer questions re the IBD BioResource), they can be recruited at the time of clinic attendance. For many participants this will be the most convenient route, and the blood sample can be taken at the time of routine clinic bloods.

- Patients will, of course, be free to take the information away with them if they wish to consider things further. If they subsequently decide they are willing to participate they can then either join up at the time of their next clinic appointment; or they can contact their local IBD team and arrange an appointment.
- We anticipate that many of our existing UKIBDGC recruits, having been notified of the IBD BioResource in the mail or by email and provided with the patient information sheet, will be willing to provide written consent through the post. Many of these individuals have been attending our clinics for regular review and management for years, and the vast majority are positively enthusiastic about research that might help others with Crohn's disease and ulcerative colitis in the future. We already hold extensive phenotype and genotype data on these individuals which, once they have consented, can be used to populate the IBD/NIHR BioResource database. This will allow the IBD/NIHR BioResource to rapidly establish a sizable cohort of patients recallable for Stage 2 studies. Where additional blood samples are required these can be obtained at subsequent clinic visits.
- For certain sites who have large cohorts of UKIBDGC patients and for whom existing holdings of frozen blood or DNA samples and data are present, these may be transferred to the IBD BioResource (following informed consent), without the need for retaking samples. A portion of thawed / extracted DNA may be returned to these sites, if required.
- Others, hearing about the IBD BioResource through another means (such as through word of mouth) will be able to submit an 'Expression of Interest' form and send this to their local IBD team. They will then be sent a formal letter of invitation and the patient information sheet, and make an appointment with a research nurse to answer questions and then, if

willing, sign consent, fill in data collection forms and provide a blood sample.

- Patients for the Inception cohort will be ascertained through a variety of routes which may vary between hospitals according to how services are configured. For example through triage of clinic referrals; in the out-patient clinic itself; patients admitted with symptomatic flares or known or suspected IBD; or at the time of endoscopy. Patients who are included on the basis of suspected IBD but who turn out on further investigation to not have IBD can remain in the Bioresource as healthy controls.

Informed Consent process: Participants will sign up to the study after being provided with the PIS and having had the opportunity to ask questions. Patients will be given the option to complete a paper or electronic consent form as provided during their clinic attendance.

5.3 Data collection and storage

Participants will be asked to complete a subject questionnaire either on paper (which they may complete in clinic or take home with them and return in a postage pre-paid envelope) or via a secure, personalised, online data collection form, a link to which will be contained within a welcome email. This will include questions relating to their IBD as well as more general fields relating to the NIHR BioResource. In addition to the health and lifestyle questionnaire, participants will also be invited to complete follow up surveys concerning surgery and medication, on a yearly basis via email, text message with URL link or letter to ensure that the IBD Bioresource database is up to date. Protocolised clinical phenotype data will be extracted from the hospital record by the clinician or research nurse and entered onto an encrypted system developed through the NIHR BioResource (the NIHR BioResource IT development team has developed the 'front end' and the back-end will comprise a MySQL server which will be hosted in the NHS environment by Cambridge University Hospital, NHS Foundation Trust and will regularly 'dump' selected pseudo-anonymised data to the NIHR Bioresource database). Core fields will include IBD type, location and behaviour, complications, co-morbidities, family history, smoking history, surgical data and drug therapy outcomes. In addition to these IBD related core fields, other more general fields will be included so that this can be incorporated into the NIHR BioResource database allowing for these patients to be re-called for any type of research studies. For the inception cohort there will be protocolised collection of follow-up data relating to disease activity, treatment

response and development of complications. In order to facilitate this process, we have developed an SMS/email messaging system and following recruitment patients will be sent, on a monthly basis, an automated message to assess if they have reached the remission/flare follow-up time-point. This will be managed centrally and the local Research Nurse will be informed if it appears that a patient is experiencing a remission/flare, so that a follow-up visit can be arranged.

All data will ultimately be uploaded to the secured NIHR BioResource database. From here it will be used for two distinct purposes:

- to enable patient selection and invitation to ethically approved Stage 2 research studies (for example where an investigator needs to access a homogenous group of Crohn's disease patients with a specific sub-phenotype). Invitation to such stage 2 studies will be carried out by the NIHR BioResource.

and

- to feed (in an anonymised form) the phenotype and treatment response data to the analysis of, for example, DNA and microbial sequencing data - to identify relevant associations and correlations.

All genotype and sequence data is stored in anonymised form at the Wellcome Trust Sanger Institute in their secure compute farm. The farm currently features >6,000 high performance compute nodes, and 1.5 petabytes of high speed lustre storage for analysis of human genetics data. The NIHR BioResource and Wellcome Trust Sanger Institute IT teams are working together to enable efficient and secure interrogation by the NIHR BioResource team of IBD BioResource genotypes without having to physically duplicate the huge existing data repository with all the attendant costs and risks which that would entail (for our IBD cohort we currently hold GWAS data with 1,000 Genomes imputation at >8,000,000 SNPs on ~15,000 subjects and whole genome sequence data on ~4,000 patients). IBD BioResource database and all previous UKIBDGC GWAS and iChip data will be deposited to the European Genome-Phenome Archive (EGA).

5.4 Samples:

All subjects in the IBD BioResource will be asked to provide blood samples for DNA and serum/plasma extraction (in the region of 20mls). The subjects in the inception cohort will in addition be asked to provide stool samples (for microbiota analysis), additional blood samples (for RNA extraction), and, if they

are undergoing colonoscopy as part of their clinical care, additional biopsies for 'omics' analyses. There is a possibility that additional samples may need to be taken if the original sample is depleted to the point that it is not possible to extract DNA. This sampling activity will all be coordinated by the research nurses at each recruitment site.

All samples will be taken and labelled with a unique barcode for pseudo-anonymisation. Blood samples for DNA and serum extraction will be submitted either to the local NIHR BioResource (where the hospital is affiliated with such a centre) – or to the National Biosample Centre (NBC) at Milton Keynes or an accredited 3rd party lab, from where they will be forwarded (after processing) to NBC. Samples will be processed using established SOP's for DNA and serum extraction. Genotyping or sequencing of the DNA samples will be required for downstream genetic association studies run by UKIBDGC or prior to subsequent Stage 2 research studies led by other investigators. Stage 2 studies involving participant recall have to be previously ethically approved and reviewed and approved by the NIHR BioResource Steering Committee.

For the inception cohort stool samples (in a postage-compatible preserving buffer kit such as Omni-gene GUT), and biopsy samples (in RNA later) will be submitted in anonymised form via Royal Mail. Processing of these samples requires substantial expertise and must be done in a single lab to avoid batch effects. UK samples will be centralized for extraction of DNA and RNA from the stool and biopsy samples, with RNA converted to cDNA and all samples barcoded and stored at -80 in a dedicated freezer prior to analysis using meta-genomic and 16s ribosomal RNA sequencing.

5.5 Reporting of results

We are planning to undertake a programme of whole genome sequencing using the samples in the IBD BioResource, to complete our interrogation of the contribution of common, rare and structural variants to IBD susceptibility. Where necessary (for example, pending the availability of the sequence data) additional genotyping may be done prior to a Stage 2 research study to identify potential participants to take part in the subsequent study. It is not planned to routinely feedback the results from genetic or other biochemical tests obtained from donated samples, but note comments above (section 3.3) regarding detection and handling of further genetic findings that may be medically actionable. Specifically, the potential clinical utility of pharmacogenetic variants known to correlate with IBD treatment outcomes in terms of efficacy and adverse events will be studied, and where 'incidental' findings are made of medically significant, treatable conditions then – where the subject has opted

for feedback - the supervising medical team will be contacted by the IBD Bioresource team with the suggestion to arrange formal testing through an accredited NHS diagnostics laboratory with subsequent relay of results by an experienced clinician or genetics counsellor.

5.6 Participant safety

During the establishment of the IBD BioResource, there is minimal potential for harm to volunteers as blood draw or (for the inception cohort) the taking of additional biopsies at the time of clinically indicated colonoscopy are the only interventions proposed and both are extremely safe. Each individual recall 'Stage 2' study will require a separate protocol and separate ethics approval.

6. The IBD BioResource funding and Organisational Structure

The IBD BioResource is funded by the Medical Research Council (see award letter attached) with additional support from the National Institute for Health Research (NIHR), the Wellcome Trust and Crohn's and Colitis UK.

The IBD BioResource will sit within the NIHR BioResource. It will have its own management committee, meeting two weekly and co-chaired by Dr Miles Parkes (Consultant Gastroenterologist in Cambridge) and Dr John Mansfield (Consultant Gastroenterologist in Newcastle).

We will maintain a close working relationship with the NIHR BioResource central team based in Cambridge (i.e. the NIHR BioResource National Coordinating Centre). The NIHR BioResource is a federation of currently thirteen local NIHR BioResources, including the NIHR BioResource Centre Cambridge, and has its headquarters (i.e. National Coordinating Centre) in Cambridge. Each local NIHR BioResource Centre has its internal organisation, including a Scientific Advisory Board that consider Researchers' applications for local Stage 2 studies (which will require separate ethics authorisation for participant recall).

The local BioResource Centres recruit participants using the national NIHR BioResource (Research Tissue Bank - RTB) ethics and study protocol and its governance structure. Stage 2 study applications can also be submitted by researchers directly to the national NIHR BioResource (RTB) for non-local studies, where they are reviewed by the NIHR BioResource Steering Committee. A local BioResource Centre may recommend to researchers applying to it to re-apply to the central/national NIHR BioResource for their Stage 2 proposed study, where it may not be feasible for that study to be run locally (e.g. if the local cohort with the inclusion criteria in question is too small).

IBD clinics based in the same institutions as one of the thirteen current NIHR BioResource centres may, where appropriate, feed their samples into their local NIHR BioResource facility, and these samples may be processed and stored locally ; the remainder of the country will post their samples to an accredited 3rd party lab . Ultimately all samples will be moved to an approved NIHR Biosample Centre.

Although it is not directly within the remit of this particular application, we summarise below how the BioResource is organised in relation to future studies:

7. Stage 2 activity-Research studies involving BioResource volunteers and/or their data

The IBD BioResource panel is open to access from researchers with ethically approved research studies. Any researcher wishing to utilise the BioResource will complete an application form indicating the phenotype or gene variant(s) of interest, it's/their likely frequency and information about what is known about its/their functional significance and association with disease. They will also summarise the nature of the experiment they propose.

These stage 2 studies may involve a range of possible options – for example

- access to specific sequence data on a specific subset of samples for detailed analysis without the need to recall participants or collection of additional clinical material
- the participation of patients recalled by genotype and/or phenotype for extra sampling e.g. the taking of one or more blood or stool samples,
- more intensive physical examinations or analyses through to potentially clinical interventions. The latter will require the study leads to set up separate clinical trials agreements with the home hospital of the BioResource volunteer (many of IBD BioResource recruiting centres have extensive experience and active clinical trials programmes).

The NIHR BioResource team will handle the process of recruitment of individuals to each new Stage 2 study.

For studies requiring recall of patients by genotype and/or phenotype, in accordance with the study requirements sequencing and or genotyping or serum sample assay will be carried out by the NIHR BioResource on all or the relevant subset of IBD samples to identify those subjects meeting genotype- or biomarker- specific inclusion criteria. Having been identified samples carrying the relevant genotypes (usually those who are homozygous for risk genotypes

and those homozygous for non-risk at specific locus of interest) or serotypes a collective list of barcodes will be passed to the BioResource volunteer administration team (who will not link e.g. risk / non-risk genotypes to identifiers). These potential volunteers will be invited to participate in the specific Stage 2 study for which they have been identified as carrying a relevant genotypes or serological markers by way of an invitation letter or email outlining details of the study. They will also be sent the approved PIS for the Stage 2 study together any other relevant study information (e.g. questionnaires). It is entirely up to them whether they wish to participate in any given Stage 2 study. The BioResource team, by controlling the process of recruitment, will be in a position to regulate the number of invitations any individual receives and the number of recall studies that volunteers participate in each year.

Patients who are willing to participate in a particular Stage 2 study for which they are eligible and to which they have been invited will contact the NIHR BioResource team and an appointment will be arranged for them to attend for sampling. To protect volunteer anonymity where the study design does not require the researcher to have direct contact with the participants and the recall can be facilitated by the BioResource, all blood sampling will be carried out by trained nurses or phlebotomists working under the direction of the NIHR BioResource. This will happen in venues appropriate for venepuncture (usually an NIHR BioResource or a local Clinical Research Facility). Informed consent will be always be taken prior to further blood sampling for Stage 2 studies – using the study-specific consent form and information sheet. Copies will be kept with volunteers' records by the BioResource team with an anonymised proof of consent form being passed on to the researcher for their records to show evidence that consent has been obtained for each sample they receive.

The researcher will conduct the study blind to genotype and only receive genotype information at the end of the study when they have completed all measurements. The genotype to phenotype link will be made without individual identifiers. In this way, individual identity and genotype will be kept separate. Also the blinding ensures that there is no bias, either in the genotyping or in the functional study.

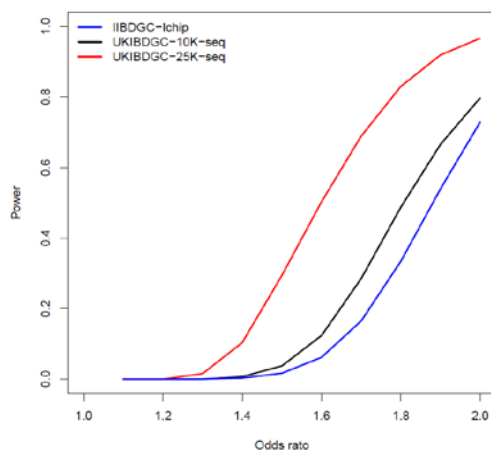
Anonymity of IBD BioResource subjects will be maintained where possible and the approved Stage 2 study design allows for it; but e.g. for intervention studies it will be necessary for the team directly undertaking the research to meet with IBD BioResource subjects. As per all participant-facing Stage 2 studies all such work will require its own ethics application and approval; and all such work will be undertaken by physicians working to GCP principles.

8 Appendix: detailed rationale for sample size

In order to ensure sufficient numbers of participants homozygous for low frequency alleles or possessing combinations of alleles can be recruited into any future stage 2 study, the size of the BioResource is crucial. We estimate that it is realistic to recruit a minimum 25,000 individuals and that this will provide adequate power for a broad range of Stage 2 studies.

This size of BioResource will benefit a range of future studies and applications that we anticipate for the cohort. Examples of these are given below.

1. through increased sample size and sampling from non-European origin groups aid fine mapping and identification of causal variants in 163 confirmed IBD GWAS loci - work by UKIBDGC + partners in the International IBD Genetics Consortium
2. significantly increase power for follow-up of low-frequency and rare variants identified by our current IBD whole genome sequencing programme at the Sanger Institute (Barrett and Anderson with UKIBDGC)



Increase in power from (current) 10,000 to (proposed) 25,000 IBD DNAs – for 1% MAF variants identified by WGS in 5,000 IBD cases, to achieve $p < 5 \times 10^{-8}$ in follow-up genotyping

3. provide expanded cohorts of IBD patients for testing and validation of markers identified in current UKIBDGC pharmacogenetic discovery programme and allow for predicting therapy outcome
4. allow selection of IBD patients by genotype or specific IBD phenotype for study of IBD pathophysiology using primary human cells. The functional effects of IBD-associated variants and their environmental modifiers will be further characterized, including on pathways relating to autophagy, ER stress, microbial sensing and IL23 response. The importance of studying primary human cells of the appropriate subset is emphasized by the cell-type specificity of eQTL effects (and indeed the

prevalence of inverse eQTL effects in different cell types) [14]. Several investigators allied to UKIBDGC are undertaking such work including: Simmons/ Powrie, Oxford; Lamb/Mansfield, Newcastle; Lee/Kaser, Cambridge

5. provide cohorts of patients for further development and validation of biomarkers of disease course / stratification based on GWAS, transcriptomic (see inception cohort below) and proteomic analysis. Gain biological insights from their functional interrogation in primary human cells; and lead into biomarker stratified intervention studies.

This would build on existing work by Lee/Parkes/Smith, Cambridge [15] and provide a validation cohort for the many stratification/biomarker studies planned or occurring in the UK and elsewhere (including PANTS Ahmad, Exeter and Hart, Imperial; IBD-BIOM and IBD-CHARACTER Satsangi, Edinburgh)

6. enable *in vitro* testing of novel/re-purposed drugs to interrogate their effect on relevant IBD pathogenic pathways stratified by genotype. This work, undertaken in collaboration with industrial partners, will precede proof of concept clinical testing. IBD BioResource will be an important channel of communication with patients, and will aid recruitment to subsequent experimental and stratified medicine studies in Crohn's disease and ulcerative colitis.
7. provide a platform to study the role of gut microbiota in IBD pathogenesis. Distinguishing cause vs correlation is complex, given the many confounders in patients who already have intestinal inflammation. At present UKIBDGC studies here are at an early stage, comprising:

- a CORE-funded pilot project by Hold, Aberdeen and UKIBDGC which has developed SOPs and infrastructure for stool processing and 16S rRNA sequence-based microbial community analysis in genotype-stratified cases and controls
- an existing (Crohn's & Colitis Foundation of America funded) collaboration between Parkes and Virgin, Washington University, St Louis, of patients in early IBD flare for stool sampling, viral enrichment, Illumina HiSeq and bioinformatic analysis to detect viral triggers. We have published early findings [16] but need to extend the size of the sample set studied.
- a pilot study comparing microbial composition of stool with samples obtained from the ileal and colonic mucosa at colonoscopy (Hart, Imperial; Hold, Aberdeen; Jostins/ Barrett, Walker/ Parkhill, Sanger) as a prelude to future specific IBD / microbiome grant application. A future 'stage 2' study might for example target stool collection <48 hours of flare, for 16S rRNA

sequencing or metagenomic methods. This study is critical for identifying microbial triggers of IBD flares, but is impossible without cohorts of patients recruited prospectively, and primed with immediate access to research teams in the event of a flare.

8. use the inception cohort of 1,000 individuals under investigation for IBD (some of whom will be newly diagnosed with IBD) – an important component of the IBD BioResource for providing more detailed samples (including stool + whole blood for RNA, alongside DNA and serum) unconfounded by drug treatment or effects of surgery with which to identify biomarkers of disease severity and course, and microbial triggers of disease flares in individuals newly diagnosed with IBD. We aim to recruit the inception cohort in years 1-3 to allow protocolised longitudinal collection of a minimum of 2 years clinical data (including outcomes and response to therapies) and repeat sampling of stool and serum.

Given the range of potential applications and experiments, the goal of recruiting 1,000 individuals under investigation for IBD patients reflects a balance of feasibility and adequate statistical power for an initial set of studies (the successful NIHR Cambridge BioResource of >17,000 volunteers initially set the same target). Where possible we will expand the sample size further, boosting the size of important phenotypic or genotypic strata to allow well powered analyses of disease subsets. Statistical power otherwise tends to be limiting in many sub-phenotype / stratified analyses. For example in 1,000 IBD patients we would anticipate ~500 UC and ~500 CD, the latter comprising ~30% ileal (n=150), 20% colonic (n=100) and 50% ileo-colonic (n=500) – important given evidence emerging from our work in the International IBD Genetics Consortium that ileal and colonic CD have distinct pathogenesis [17].

The following are examples of existing or potential future studies, which would benefit from the availability of the inception cohort. Such an inception cohort would:

1. provide accurate prospective clinical data to enhance phenotype-genotype analyses, including response to drug therapies for pharmacogenetic analysis (UKIBDGC core). In this respect it will provide both a valuable validation panel for existing programmes and a platform for a new series of studies. In 1,000 cases we would expect
 - ~450 UC patients to be treated with mesalazine; 50% remission
 - ~400 IBD patients with thiopurine; 30% remission
 - ~200 CD patients with anti-TNF; 40% remission

Hence a study seeking a proteomic or transcriptomic biomarker of thiopurine efficacy could access inception samples from ~120 subjects who gained

remission after treatment. This would, for example, provide 95% power at $P < 0.01$ to identify a biomarker with frequency 30% correlating with an OR of 3 for remission.

2. deliver clinical data and bloods unconfounded by the effects of chronic disease or drug treatments, for epigenetic analysis (Satsangi, Edinburgh; Parkes, Cambridge); and for proteomic and immunological analyses – with the option to resample the same individuals to obtain fresh material for cell-based studies (Simmons, Oxford; Kaser, Cambridge).
3. deliver biometric data unconfounded by the effects of chronic disease or drug treatments for epidemiologic analyses (Hawkey, Nottingham);
4. deliver whole blood RNA + clinical outcomes data for development / validation of transcriptomic biomarkers (Lee/Smith, Cambridge)
5. include environmental data for genetic interaction analysis (Barrett/Anderson, Sanger)
6. provide a unique resource of stool samples to detect microbial triggers of IBD. Previous work suggests that, unlike *Helicobacter pylori* in peptic ulcer, IBD is not caused by a single microbial species. However, early studies have been underpowered to detect low abundance taxa such as are known to trigger IBD in animal models, and none have sampled large cohorts at diagnosis. Our inception cohort would be well powered and provide samples un-confounded by the effects of drugs, surgery or changes in diet; and repeat sampling at first remission and first flare would identify changes in the microbiota through periods of remission and relapse, with availability of paired samples further improving statistical power.

Assuming access to the NIHR Cambridge BioResource of healthy controls, study of our population would have 80% power to detect associations between common (>25%) microbial elements with an OR=1.4, and rare elements (<1% frequency) with OR >3, at $p < 0.01$. Through longitudinal sampling the CD inception cohort of 500 cases would provide 96% power to detect a microbe present in 2% of cases in remission rising to 10% (an OR of 5) on significant relapse at $p < 0.01$.

9. References

1. Parkes, M., et al., *Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility*. Nat Genet, 2007. **39**(7): p. 830-2.

2. Wellcome Trust Case Control, C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls*. *Nature*, 2007. **447**(7145): p. 661-78.
3. Barrett, J.C., et al., *Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease*. *Nat Genet*, 2008. **40**(8): p. 955-62.
4. UK IBD Genetics Consortium and WTCCC., *Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region*. *Nat Genet*, 2009. **41**(12): p. 1330-4.
5. Fisher, S.A., et al., *Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease*. *Nat Genet*, 2008. **40**(6): p. 710-2.
6. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci*. *Nat Genet*, 2010. **42**(12): p. 1118-25.
7. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease*. *Nature*, 2012. **491**(7422): p. 119-24.
8. Huang H et al. Fine mapping IBD loci to single variant resolution. *Nature* Jul 2017; 547 (7662); 173-178
9. Cleyneen, I., et al., *Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study*. *Lancet*, 2016. **387**(10014): p. 156-67.
10. Ellinghaus, D., et al., *Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies*. *Gastroenterology*, 2013. **145**(2): p. 339-47.
11. Lee, J.C., et al., *Human SNP links differential outcomes in inflammatory and infectious disease to a FOXO3-regulated pathway*. *Cell*, 2013. **155**(1): p. 57-69.
12. Barroso, I., et al., *Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension*. *Nature*, 1999. **402**(6764): p. 880-3.
13. Altshuler, D., et al., *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes*. *Nat Genet*, 2000. **26**(1): p. 76-80.
14. Fairfax, B.P., et al., *Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles*. *Nat Genet*, 2012. **44**(5): p. 502-10.

15. Lee, J.C., et al., *Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis*. J Clin Invest, 2011. **121**(10): p. 4170-9.
16. Norman, J.M., et al., *Disease-specific alterations in the enteric virome in inflammatory bowel disease*. Cell, 2015. **160**(3): p. 447-60.
17. Cleyne, I., et al., *Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study*. Lancet, 2016. **387**(10014): p. 156-67.